DOCUMENT RESUME

ED 141 791 CS 003 576

AUTHOR Venezky, Richard L.; And Others
TITLE On-Line Diagnosis of Reading Difficulties-II. Technical Report No. 386.
INSTITUTION Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Sep 76
CONTRACT NE-C-00-3-0065
NOTE 46p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Computer Assisted Instruction; Decoding (Reading); Primary Education; *Reading Diagnosis; *Reading Research; *Reading Skills; *Test Construction; Test Reliability
IDENTIFIERS *PLATO

ABSTRACT

Diagnostic tests for nine different letter-sound patterns were developed for the PLATO Computer Assisted Instruction system and were administered to 122 first, second, and third graders. Results for four of the patterns showed low test reliabilities and no monotonic increase for the percentage correct with increasing grade level. The remaining tests had relatively high reliabilities and scores increasing monotonically with increasing grade level. Correlations between tests were relatively low, except for the vowels in monosyllabic words. Rank-order correlations between grades, based on ranking of tests by percentage correct within grade, varied from 0.87 to 0.65. A relatively high percentage of subjects (25%) showed evidence of random responding. From these results, an improved diagnostic system has been planned, using coarse to fine grain progression and response-contingent item and test selection. (Author/AA)

Technical Report No. 386

ON-LINE DIAGNOSIS OF READING DIFFICULTIES-II

by

Richard L. Venezky, Susan Chicone, and James Perry

Report from the Project on
Conditions of School Learning and Instructional Strategies

Richard L. Venezky
Faculty Associate

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

September 1976

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the National Institute of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by that agency should be inferred.

Center Contract No. NE-C-00-3-0065

WISCONSIN RESEARCH AND DEVELOPMENT CENTER FOR COGNITIVE LEARNING

MISSION

The mission of the Wisconsin Research and Development Center for Cognitive Learning is to help learners develop as rapidly and effectively as possible their potential as human beings and as contributing members of society. The R&D Center is striving to fulfill this goal by

- conducting research to discover more about how children learn
- developing improved instructional strategies, processes and materials for school administrators, teachers, and children, and
- offering assistance to educators and citizens which will help transfer the outcomes of research and development into practice

PROGRAM

The activities of the Wisconsin R&D Center are organized around one unifying theme, Individually Guided Education.

FUNDING

The Wisconsin R&D Center is supported with funds from the National Institute of Education; the Bureau of Education for the Handicapped, U.S. Office of Education; and the University of Wisconsin.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Diagnostic tests for 9 different letter-sound patterns were developed for the PLATO CAI system and administered to 122 first, second, and third graders. The purpose of the tests was to assess the feasibility of on-line diagnosis of basic reading skills. Results for four of the patterns--initial g, final g, final c, and bisyllabic a--showed low test reliabilities and no monotonic increase for percentage correct with increasing grade level. The remaining tests had relatively high reliabilities and monotonically increasing scores with increasing grade level. The most reliable subtests for separating grade level were the soft pronunciation of c and the long and short pronunciations of i, o, and u.

Correlations between tests (uncorrected for attenuation) were relatively low, except for the vowels in monosyllabic words, which varied from .57(i-o) to .30(i-u). Rank order correlations between grades based on ranking within grade of tests by percentage correct varied from .87 (grades 2 and 3) to .65 (grades 1 and 3). A relatively high percentage of the subjects (approximately 25%) showed evidence of random responding.

From these results, an improved diagnostic system has been planned, using coarse to fine grain progression and response contingent item and test selection.

I

INTRODUCTION

The rationale for an individualized approach to education is the need to offer each child the best opportunity to realize fully his or her own potential. Individualization, however, is not often easily implemented. An individualized approach to teaching reading, for example, must be preceded by a thorough, accurate assessment of each child's reading abilities. The complexity of the reading task presents difficulties in arriving at this assessment, for reading is not one skill, but rather a synthesis of many skills. Such subskills as literal comprehension, word recognition, and letter-sound generalizations are all part of the initial reading process, and each should be assessed in an individualized, skill approach to reading instruction. But without one or more aides, the average classroom teacher cannot be expected to do extensive testing of these (or any other) skills. To assess, score, and record the results for each of 25-30 children in each of 5-10 reading skills requires far more time than most teachers are willing or able to spend on assessment.

An alternative to teacher or reading specialist testing is on-line computer testing whereby the student receives test items at a computer terminal and responds by typing on a keyboard or touching designated areas on a screen.

Using a computer for on-line reading skill assessment has obvious and distinct advantages. It can, within the range of response types which it can handle, be faster, more extensive, and more accurate than human assessment. Through the use of response contingent testing, children can be diverted from tasks which are inappropriate for them. And finally, by programming data analysis and report generation as part of the testing system, assessment results can be available almost immediately after testing is completed, without a need for hand scoring, coding, or key punching.

On-line testing also has distinct disadvantages, especially for younger children, in that oral responses cannot be evaluated and the degree of intelligence which the terminal can display is limited by practical considerations of program size and complexity. Nevertheless, the potential advantages of on-line reading skill diagnosis are attractive enough to merit a feasibility study.

To this end a PLATO terminal, consisting of a plasma display panel, touch-sensitive collar, rear projection microfiche unit, audio device, and printer has been utilized over the past two years in pilot studies of on-line skill diagnosis. This system was selected over a variety of other available computer-assisted instruction (CAI) systems

primarily because of its display capabilities and its software support. Systems which utilize teletypes or alphanumeric displays lack the graphics capability which is essential for testing primary level children, while most other graphic terminals lack software especially designed for CAI applications.

The first phase in the development of a series of on-line diagnostic tests was to establish the PLATO system's validity as a testing instrument. To that end, two groups of fifth and sixth grade children were given a standardized reading comprehension test in April of 1974; half took the test on the PLATO terminal and half took the test using paper and pencil. It was found that, given similar testing conditions, the results obtained from an on-line presentation were not significantly different from those obtained with paper and pencil testing (Venezky, Bernard, Chicone, & Leslie, 1975).

Work was then started on the development of tests for reading subskill abilities, beginning with tests for letter-sound generalizations. This skill area was selected because of its importance for the acquisition of reading ability, and because the present project staff had already developed testing procedures and compiled developmental data on the acquisition of letter-sound patterns (Calfee, Venezky, & Chapman, 1969; Venezky & Johnson, 1973; Venezky, 1974).

II

EXPERIMENTAL STUDIES

PRELIMINARY STUDIES

The on-line procedures and test items for measuring letter-sound generalizations were developed by trial and error, beginning with adaptations of procedures used in earlier studies. The stimulus words had to be synthetic to ensure that letter-sound generalizations and not sight word recognition would be measured. Some of the words, particularly some long and short "a" words and hard and soft "c" and "g" words had already been tried out in previous oral-response studies (Venezky, Chapman, & Calfee, 1972). New words were generated to enlarge stimulus pools or to test new letter-sound patterns.

Pilot tests revealed that the wording of the directions for the test as well as the method of presentation had a major effect on children's performances. Two problems were revealed by these preliminary studies. First, some children gave real word responses for most of the items. This problem was corrected by eliminating real words that had been included in the practice items and by using the descriptors "made up" and "brand new" when talking to the children about the words they were going to see. Another modification of the directions, seen more as a necessity to make the test self-contained and consistent, was the transition from proctor introduction to machine introduction. In the initial tests a proctor explained most of the directions individually to each child, with the result that there was variation in the directions which children received, particularly in relation to practice item responses. The final version of the test was given entirely through a recording on the audio disk. Children were guided through several warmup items and then into the test solely through the recording. Feedback varied in amount and kind according to the individual child's performance on the test items, so that each child's understanding of the task was still ensured.

The method of presentation for test words was a primary concern on the pilot tests. Two testing formats were developed and tried out before a third was finally settled upon. The first of these formats was a same/different paradigm in which the child viewed a word on the screen while he or she heard an animal pictured on the screen give the correct or a variant pronunciation. The child then touched a "yes" or "no" box on the screen depending on whether he or she agreed with the pronunciation or not. The second format was a triple choice paradigm. For this test the child was asked to be a judge. Three different animals gave pronunciations for the word shown (correct, plausible, and

implausible) and the child was asked to touch the animal that said it the "right way." For each test the child could hear the pronunciations and change his or her answer as many times as desired. Both the same/different format and the triple-choice format had the drawback that children were given pronunciations for the words rather than having to generate them on their own.

The format ultimately decided upon was a two-choice sorting game which incorporated all the revisions made as a result of previous tryouts of the words and the directions. This format was designed so that children would have to generate the sound for a specified letter in a synthetic word and then select the picture whose name began with that same sound. This format was pilot tested on nine second and third grade children. As a result of the pilot test the directions were expanded and more patterns were added for the final version.

In addition, a variety of analysis and reporting routines were developed on the PLATO system so that test results could be presented soon after the last subject was tested. The following sections describe the results of administering this test to first, second, and third grade children and show some of the capabilities of the analysis and display routines. A description of the computer programs is given in Appendix A.

III

<u>METHOD</u>

STIMULUS ITEMS

For each of 9 letter-sound patterns, 12-16 synthetic words were generated. The patterns selected were:

1. initial <u>c</u> (/k/ or /s/)
2. final <u>c</u> (/k/ or /s/)
3. initial <u>g</u> (/g/ or /ǰ/
4. final <u>g</u> (/g/ or /ǰ/)

5-8. the long and short pronunciations of <u>a</u>, <u>i</u>, <u>o</u>, and <u>u</u> in monosyllabic words

9. the long and short pronunciations of <u>a</u>, preceding single and double consonants in two-syllable words (e.g., <u>tappen</u> vs. <u>tapen</u>)

Most of the words, with the exception of those in the last category, were three to five letters in length. (See Table 1 for the complete list of stimuli.)

PROCEDURES

<u>Testing Method</u>

Test items were administered on PLATO terminals which were being used in the testing classrooms for reading instruction during the year. The number of terminals in each classroom varied from one to three. Two project staff members, each at a different school, supervised the testing.

Each child worked one-to-one with a PLATO terminal. Instructions for the test had been previously recorded on audio disks so that, except for a brief introduction to the machinery by the staff members (which some children did not need) the children worked independently, receiving all instructions through earphones.

The testing session began with five warmup items. During the course of the warmup items children were familiarized with the test format and introduced to the procedures for recording and changing

TABLE 1

STIMULI FOR MAJOR PATTERNS

Warmup items

feep
boop
flope
pode
dop

Initial c (kangaroo, seal)

cib
cerp
cofe
cilf
cobe
cempt
cung
cyfe
cabe
cefe
cest
carg
cylm
cuse
cipe
cirt
cym
cose
cade
corb

Initial g (girl, giraffe)

gest
geft
gept
gim
gite
geme
gind
gink
gube
gurk
gon
gope
gade
gabe
golb
gand

Monosyllabic a (cake, cat)

cale
dack
lase
tasp
dape
prane
bab
vabe
fam
nad
tafe
tand
plab
tass
plafe
bap
hade
lan
slape
dap
clade
nast
calse
trabe

Final g (cage, frog)

noge
pog
bage
lig
tuge
sug
darg
porge
sege
mige
reg
dag

Final c (hook, glass)

woc
dac
jic
roce
bec
nuc
hac
fice
mic
tuce
poce
mece

Bisyllabic a (cake, cat)

nattis
dallor
nammis
lapple
cannod
sattom
natis
dalor
namis
laple
canod
satom

Monosyllabic i (kite, gift)

jik
slin
kip
bim
jit
sib
jike
sline
kipe
bime
jite
sibe

Monosyllabic o (boot, sock)

jod
mot
wom
tob
rog
lop
mote
wome
tobe
roge
lope
jode

Monosyllabic u (boat, tub)

fub
pud
vun
huk
slub
stut
fube
pude
vune
huke
slube
stute

answers. They also learned how to have the instructions repeated. Negative or positive feedback, depending on children's responses to the items, was given through the audio unit for all sample items. Children were allowed as much time as they needed to sequence themselves through the sample test.

Testing Format

For each of the nine letter-sound patterns, two pictures (e.g., a giraffe and a girl) appeared in separate houses at the top of the screen. (See Table 1 for a listing of all the pairs of pictures shown.) Children were told for this set of pictures:

"Look at the two pictures at the top of the screen.

One picture is a giraffe. One picture is a girl." As each picture was introduced by the audio unit, it was blinked several times to attract the subject's attention. Then one of the stimulus words (e.g., gim) appeared in a box at the center of the screen. The letter that the children were to attend to (g) was underlined. Then, a second audio message stated:

"Read the word. Then touch the picture that has the same sound as the underlined letter." Children could hear these instructions as well as the picture names again by touching a circle at the bottom of the screen.

As soon as a child indicated his or her choice by touching one of the houses, the word appeared in that house. The child was then to touch the same house again and a new word would be shown to him. The double touch was instituted in prior studies (see Venezky, Bernard, Chicone, & Leslie, 1975) to avoid mistaken touching. No reinforcement was given on test items. After receiving six items for a pattern, children were presented with two new pictures and six new words testing a different letter-sound pattern. This procedure was repeated until all nine patterns were completed. Upon completion of the test, a message came on the screen signaling the end of the session. The average time for warmup was 4 minutes, 42 seconds; for testing (54 items), 11 minutes, 55 seconds--or about 13.2 seconds per item.

SUBJECTS

Forty-one first, 43 second, and 38 third grade children from public schools in Champaign-Urbana, Illinois, participated in the study. The mean ages were 6.8 years, 8.0 years, and 9.0 years respectively. The children attending the Champaign schools were using the PKR (Phonetic Keys to Reading) program while the children attending the Urbana school were using the Holt-Rinehart program. About half of the children had previous PLATO experience. The sample was drawn from a fairly wide range of socio-economic backgrounds.

IV

RESULTS

PATTERNS

The percentages correct for each grade and for sex (collapsed across grade) are shown in Table 2. Sex differences tended to be small, with males doing slightly better than females. Ability to select the proper generalization increases or holds steady with increasing grade for all patterns except initial g, final g, final c, and bisyllabic a.

TABLE 2

PERCENTAGES CORRECT BY GRADE AND SEX FOR EACH TEST

	Grade			Sex		Total
	1	2	3	M	F	
1. initial c	48	51	59	55	50	53
2. initial g	50	47	49	49	49	49
3. a-mono	58	67	66	66	61	64
4. final g	50	65	59	58	58	58
5. final c	55	53	60	55	57	56
6. a-bisyll	50	54	50	55	47	51
7. i-mono	65	69	74	70	68	69
8. o-mono	66	71	74	72	68	70
9. u-mono	55	63	67	61	62	61

A division of each pattern into two subpatterns (long versus short for vowels; hard versus soft for c and g) shows more clearly the generalizations which are (or are not) developing across the three grades which were tested (Table 3).

c-initial

For c in environments in which it should be /k/ as in coal, the percentage of correct responses starts fairly high (63%), but shows almost no growth from grade 1 to grade 3; selection of the correct response for c when it should be /s/ as in city increases, however, from 36% in grade 1 to 54% in grade 3.

g-initial

Soft g shows a similar increase to that of soft c, but third graders are still selecting only 42% correctly. Hard g, on the other hand, falls from 70% correct in grade 1 to 56% correct in grade 3.

c-final

c before final e shows an erratic developmental pattern, although final c develops from 55% correct in grade 1 to 66% correct in grade 3.

g-final

The g-final patterns are nearly identical to the c-final patterns, both in development trends and in third grade percentages correct.

bisyllabic a

Both long and short bisyllabic a show erratic developmental patterns. Short a responses start at 60% correct in grade 1 and decrease to 56% correct in grade 3. Long a responses increase from grade 1 to grade 2, but decrease from grade 2 to grade 3

monosyllabic vowels (a, i, o, u)

With some minor exceptions, the long and short monosyllabic vowels show a smooth increase in percentages correct across grade levels, and relatively high percentages of correct responses by third grade.

TEST ITEMS

Percentages correct for each test item, organized by pattern, are shown in Appendix B. The ranges for each pattern are summarized in Table 4.

TABLE 3

PERCENTAGES CORRECT BY GRADE,
SEX, AND CLASSROOM
FOR EACH SUBTEST

	Grade			Sex		Classroom						
	1	2	3	M	F	wv1	wl1	wl2	co2	wv3	co3	Total
1. c-initial/soft	36	44	54	51	37	42	32	52	33	56	52	44
c-initial/hard	63	63	65	60	67	68	57	60	65	64	66	63
2. g-initial/soft	30	39	42	37	38	37	24	33	47	32	52	37
g-initial/hard	70	55	56	62	59	67	72	57	54	62	50	60
3. a-mono/short	68	71	69	74	65	64	70	77	65	83	58	69
a-mono/long	49	63	62	59	57	58	40	59	68	61	64	58
4. ge-final/soft	43	63	52	53	52	48	38	63	62	44	60	52
g-final/hard	59	67	65	64	63	54	64	64	69	70	61	64

(continued)

TABLE 3 (cont.)

	Grade			Sex		Classroom						
	1	2	3	M	F	wv1	wl1	wl2	co2	wv3	co3	Total
l/soft	55	49	52	51	53	64	48	44	53	58	46	52
/hard	55	57	66	58	60	66	46	54	61	77	56	59
l/short	60	58	56	63	54	64	56	63	53	57	55	58
l/long	39	50	45	48	42	51	29	49	52	44	47	45
short	59	70	78	69	69	64	55	75	64	85	73	69
long	70	68	69	71	67	69	72	76	59	79	59	69
short	74	74	77	79	71	77	72	80	69	79	75	75
long	57	66	71	65	64	64	52	63	71	73	70	65
short	66	73	75	66	77	65	66	79	66	81	71	71
long	44	53	59	55	47	53	36	55	51	63	55	52

TABLE 4

PERCENTAGES CORRECT FOR HIGHEST AND LOWEST ITEMS IN EACH PATTERN

Pattern	Number of Items	Percentage Correct	
		Highest	Lowest
c-initial/soft	11	60	30
c-initial/hard	9	76	53
g-initial/soft	8	49	25
g-initial/hard	8	72	54
a-mono/short	12	83	57
a-mono/long	12	68	42
ge-final/soft	6	64	36
g-final/hard	6	87	44
ce-final/soft	5	62	39
c-final/hard	7	75	49
a-bisyil/short	6	75	47
a-bisyll/hard	6	59	34

(continued)

TABLE 4 (cont.)

		Percentage Correct	
Patterns	Number of Items	Highest	Lowest
i-mono/short	6	77	63
i-mono/long	6	79	63
o-mono/short	6	81	58
o-mono/long	6	70	58
u-mono/short	6	79	61
u-mono/long	6	67	44

The monosyllabic vowel patterns tend to have the greatest homogeneity among items, although several exceptions exist, particularly for long u.

A more sensitive picture of item reliabilities can be seen in Appendix B where quartile scores are shown for each test item. The percentage of items within each pattern which either increase (in percentage of correct responses) or remain the same across each quartile from lowest to highest varies from 88% (initial g) to 50% (g-final, and monosyllabic a). Six patterns have 75% of their items in this category.

Correlations Between Tests

Correlations between tests, shown in Table 5 for all subjects, tend to be quite low, except for those among the monosyllabic vowel patterns.

TABLE 5

CORRELATIONS BETWEEN TESTS FOR ALL SUBJECTS (N=122)

	1	2	3	4	5	6	7	8	9
1. c-initial	-	.04	.07	.04	.05	.02	-.01	.11	.09
2. g-initial		-	.02	.10	-.09	-.04	.06	.06	-.02
3. a-mono.			-	.27	.13	.14	.44	.52	.45
4. g-final				-	.07	.15	.24	.33	.27
5. c-final					-	.22	.25	.27	.11
6. a-bisyll.						-	.17	.22	.18
7. i-mono.							-	.57	.30
8. o-mono.								-	.35
9. u-mono.									-

Correlations With Teacher Ratings

Teacher ratings for reading ability were correlated with combined scores for the monosyllabic vowel patterns; correlation coefficients varied from .86 (grade 2, Wiley School) to .19 (grade 3, Westview School). No pattern was discernible, however, either by grade level or school.

Correlations Across Grades

Patterns were rank ordered within grade according to difficulty; rank order correlations for these orderings were: .68 (r_{12}), .65 (r_{13}), and .87 (r_{23}).

SCORE DISTRIBUTIONS

For the 122 Ss the range of total scores (54 items) was from 20 to 49 correct. Scores were normally distributed with a peak at around 27 (which also represents the number correct for guessing).

Response Bias

An analysis for response bias was done by assigning each response to one of four quartiles, based on position of correct response (right or left) and position touched (right or left). A subject was then classified as a random responder if the minimum value in his quartiles was less than 20%. By this process 29 random responders were identified: 8 first graders (19.5%), 11 second graders (25.6%), and 10 third graders (26.3%). Only one subject showed a position bias (here defined as 80% or more responses on the same side).

V

DISCUSSION

These results are not especially easy to interpret, due to reliability problems with certain tests, and perhaps even with the two-choice sorting paradigm itself. On the one hand, the high (almost 24%) number of subjects who showed random response patterns, and their fairly even distribution across all three grade levels, makes the testing paradigm suspect. On the other hand, the relatively high rank order correlations between grades indicates consistency in test difficulties. Four of the nine test patterns were clearly beyond the abilities of even the third graders and may account for a major portion of the random responses. These were (1) initial g, (2) final g, (3) final c, and (4) bisyllabic a. Letter-sound generalization studies by Venezky, Chapman, and Calfee (1972) also found little generalization of g patterns, even with sixth graders.

The remaining tests showed an expected increase in correct responses with increasing grade level, in addition to high item reliabilities. Comparisons of the results from these tests to comparable data from Venezky and Johnson (1973), and Venezky, Chapman, and Calfee (1972) show similar results for patterns that children are beginning to master, but lack of discrimination for those on which few children showed mastery (due obviously to the two-choice procedure). Since the on-line procedure is intended only for individual mastery testing, this limitation is not serious. It should be noted, nevertheless, that the choice of alternatives in the two-choice procedure has a marked effect upon test results when the subjects have not mastered the pattern tested. For initial c before e, i, or y as an example, there is a strong response bias for /k/ among primary level children. If, therefore, /k/ is the alternative sound, selection of /s/ will be well below the 50% guessing rate. If, however, other alternatives are used, none of the selections will be preferred, and therefore the selection of the /s/ alternative will be higher than in the former case. Given this effect, results below the mastery level should not be interpreted as having any other significance than indicating non-mastery.

From the results so far, we have concluded that on-line testing of letter-sound generalizations is possible, but that several major steps need to be accomplished before such testing will be practicable. The first is to build a multi-level testing capability that uses response-contingent techniques to branch students from gross to fine level testing (see Wood, 1973). At present we envision a three level system in which the first level divides subjects into four classes: (1) those who have mastered all the patterns which the system can test;

(2) those whose generalizations are so poor or whose abilities to interact with the system are so limited that no further information can be obtained; (3) those who have clearly mastered the simplest patterns but may not have mastered the more complex ones; and (4) those who may not have mastered the simplest patterns, but who are testable.

Group 1 would be rejected out the top of level 1 and group 2 rejected out the bottom. Group 3 would be branched to the second level which would survey general decoding ability and then branch to tests of specific generalizations (e.g., stress patterns) on a third level. Group 4 would follow a similar route, but with more elementary tests.

Whether to continue using the two-choice sorting procedure, or to experiment further with alternatives has not been decided. Most testing paradigms, including the three-choice matching procedure which produced nearly acceptable results in a pilot test, require higher quality audio than is now available. Purchase of an improved audio unit will be considered, however.

Utilization of several different testing procedures is also a consideration but is probably not acceptable because of the time required to explain the various procedures to a subject. Children did not exhibit fatigue or boredom with these tests, although they were required to sit for an average of about 16-1/2 minutes each. Of this time almost five minutes was absorbed in the introduction to the sorting procedure. A second five minute explanation might raise the session time to about the tolerance level for primary level children, and thereby limit the total number of items which could be tested.

Finally, a verbal reporting system needs to be developed so that information interpretable by the classroom teacher can be generated for each child tested. This will be pursued as soon as the tri-level testing procedure is developed.

REFERENCES

Calfee, R. C., Venezky, R. L., & Chapman, R. S. Pronunciation of synthetic words with predictable and unpredictable letter-sound correspondences. Technical Report No. 71. Madison: Wisconsin Research and Development Center for Cognitive Learning, 1969.

Venezky, R. L. Theoretical and experimental bases for teaching reading. In T. A. Sebeok (Ed.), Current trends in linguistics. Vol. 12. The Hague: Mouton, 1974, Pp. 2058-2100.

Venezky, R. L., Bernard, S., Chicone, S., & Leslie, R. On-line diagnosis of reading difficulties. Technical Report No. 327. Madison: Wisconsin Research and Development Center for Cognitive Learning, 1975.

Venezky, R. L., Chapman, R. S., & Calfee, R. C. The development of letter-sound generalizations from second through sixth grade. Technical Report No. 231. Madison: Wisconsin Research and Development Center for Cognitive Learning, 1972.

Venezky, R. L., & Johnson, D. Development of two letter-sound patterns in grades one through three. Journal of Educational Psychology, 1973, 64, 109-115.

Wood, R. Response-contingent testing. Review of Educational Research, 1973, 43, 529-44.

APPENDIX A

PLATO/TUTOR ROUTINES FOR TEST
CONSTRUCTION, TEST ADMINISTRATION, AND DATA ANALYSIS

TEST DESCRIPTION AND ADMINISTRATION

Four major programs have been developed for use in the study of letter-sound generalizations. Two of these programs are general purpose programs and could likely be used with other studies with little or no modification. One of the general purpose programs is a test editor which can be used for entering and modifying test words and their associated codes. Ten test blocks can be entered and edited, each containing up to 30 test words, 60 three-character test word codes, and 90 ten-character response codes, which describe the test word and alternate responses. Each test block may also contain a 60-character title. For an overview of the common storage, see Figure A-1; for details of the test block see Figure A-2.

The second general purpose program is used for entering data on the test subjects. This consists of name (18 characters), sex (1 character), age (1 character), school code (2 characters), class code (1 character), and group code (2 characters). In addition, 300 two-bit bytes arranged in ten 30-byte variables are available for storing item responses. These correspond to the ten 30-word test blocks the test editor is capable of editing. The data in this area can be displayed on the plasma panel, printed, or passed as input to analysis routines. There is also a unit for printing the completed roster of subjects.

The two other programs are more specialized; parts of them could be used in other studies, but only after some modification. The first is the wordsort test driver (see Figures A-3 and A-4). The driver presents each test word, senses the subject's response, and records that response in the subject's records. The driver requires a touch-pause-touch response to each test word (where the same answer is touched both times). This prevents a subject unfamiliar with the Plato touch collar from making an accidental response.

DATA ANALYSIS

The second special purpose program contains units which perform statistical analyses and print results. These include the printing of histograms, analysis of test items, correlations between test blocks, and means and standard deviations of various data.

The unit prstudat prints out the raw (right-wrong) data with no analysis or interpretation. This was used to preserve the data from catastrophic loss, and can be done at any point in testing. Unit histdata collects the scores for all subjects and hprint prints a histogram from the data. Histdata is also used before hdisplay which displays a histogram which is used to obtain quartile distributions as decided by the operator of the lesson. The quartiles are then used in an item by quartile print generated in item-quart. Unit wditemch prints item by choice data, used in item analysis. Unit correl finds the correlations between each of the nine tests. Blockdat summarizes scores by sex, grade, and classroom, as well as certain other classifications. Unit rtbias checks each of the subjects for response bias (right or left) and for random responding on the entire set of tests. Ritebias

checks for a bias in each test. Meanage finds and prints the mean age of subjects by grade. Spscore prints the total number of right answers for four selected test blocks (3, 7, 8, 9) for each subject. Unit time finds and prints the time for warmup and the total testing time for each student; at the end of testing it prints mean times. Mtimeq calculates the mean time for selected quartile groupings.

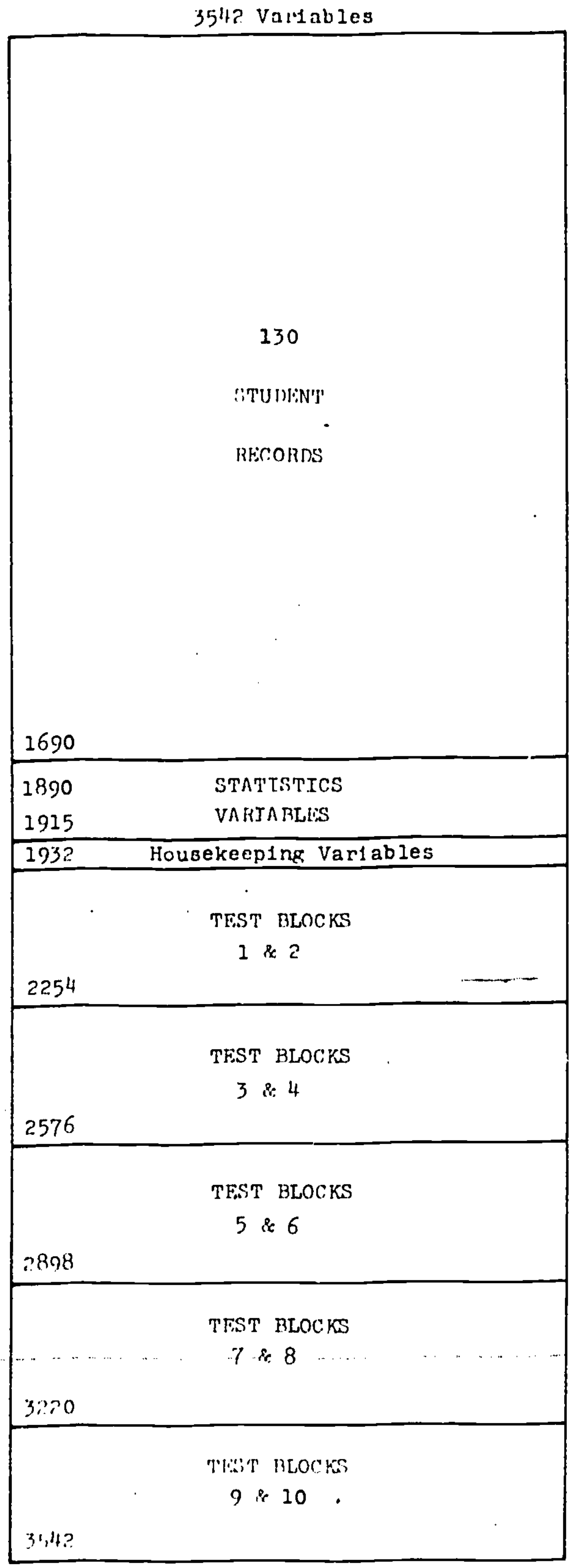


Figure A-1. Common map.

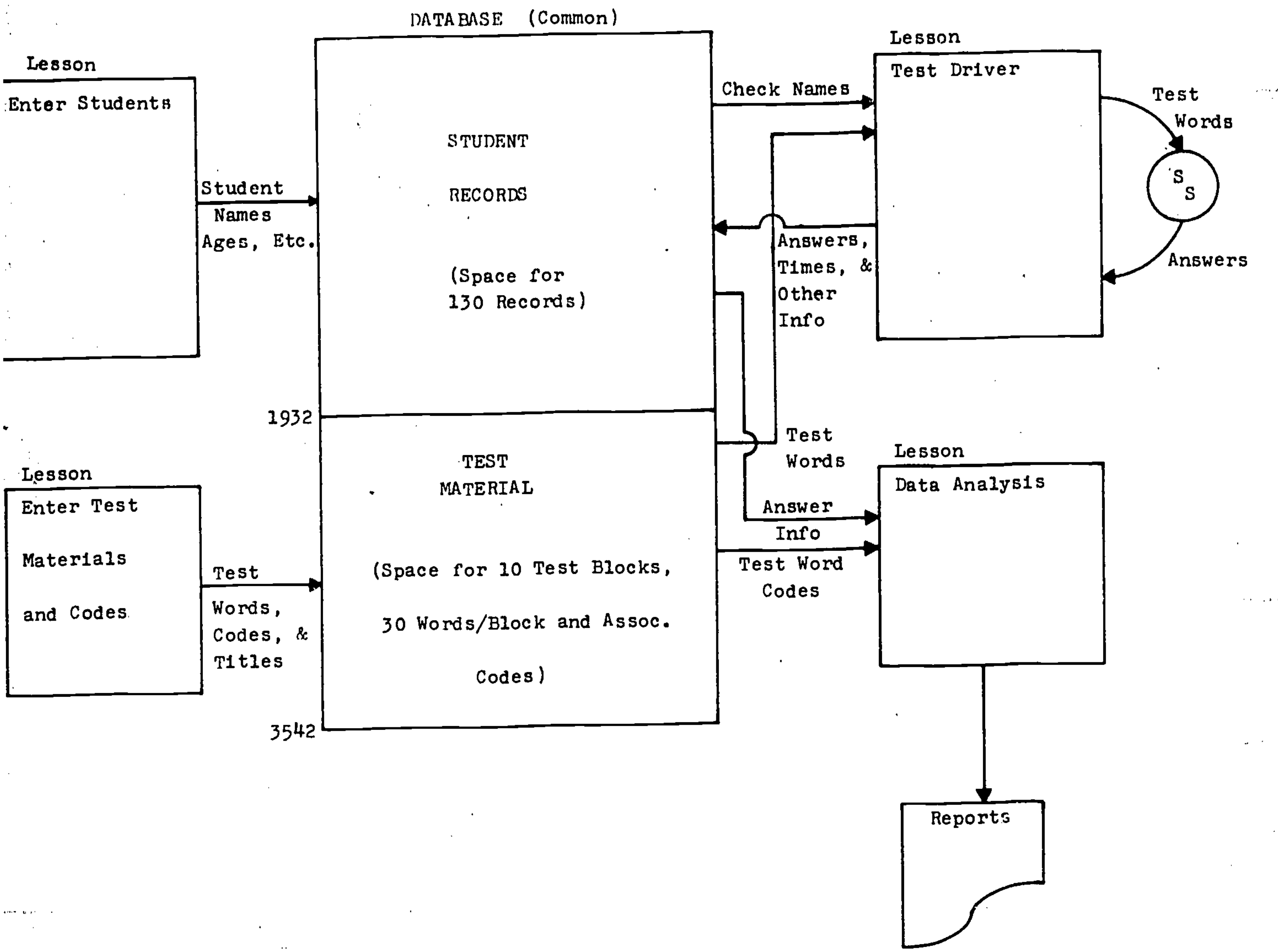
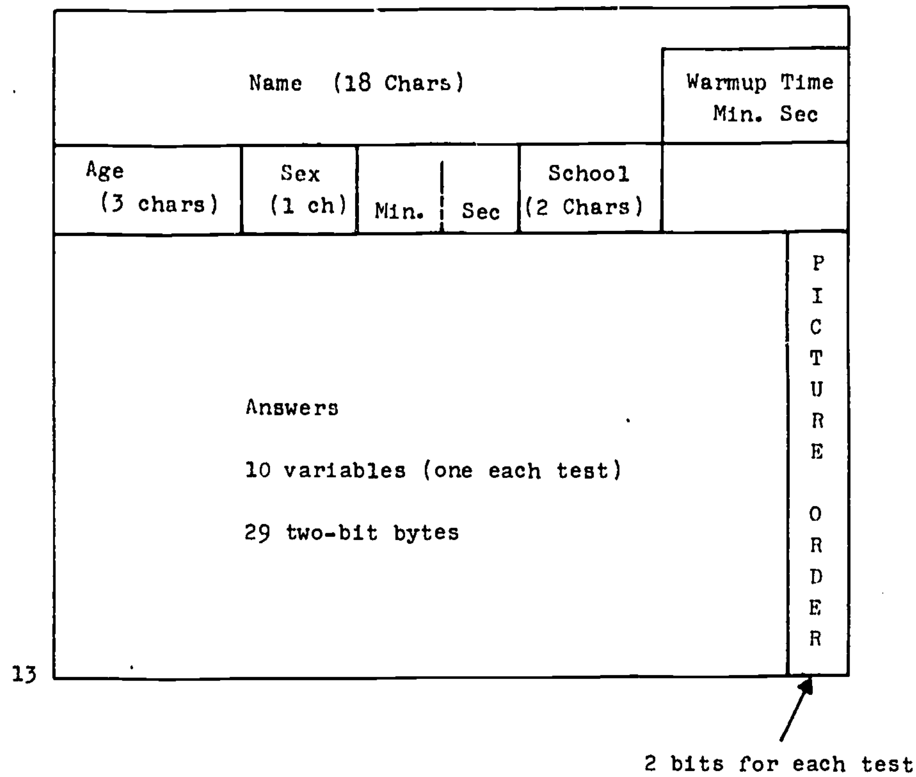


Figure A-2. Schematic diagram of total test system.

Detail of a
Student Record



Detail of a
Test Block

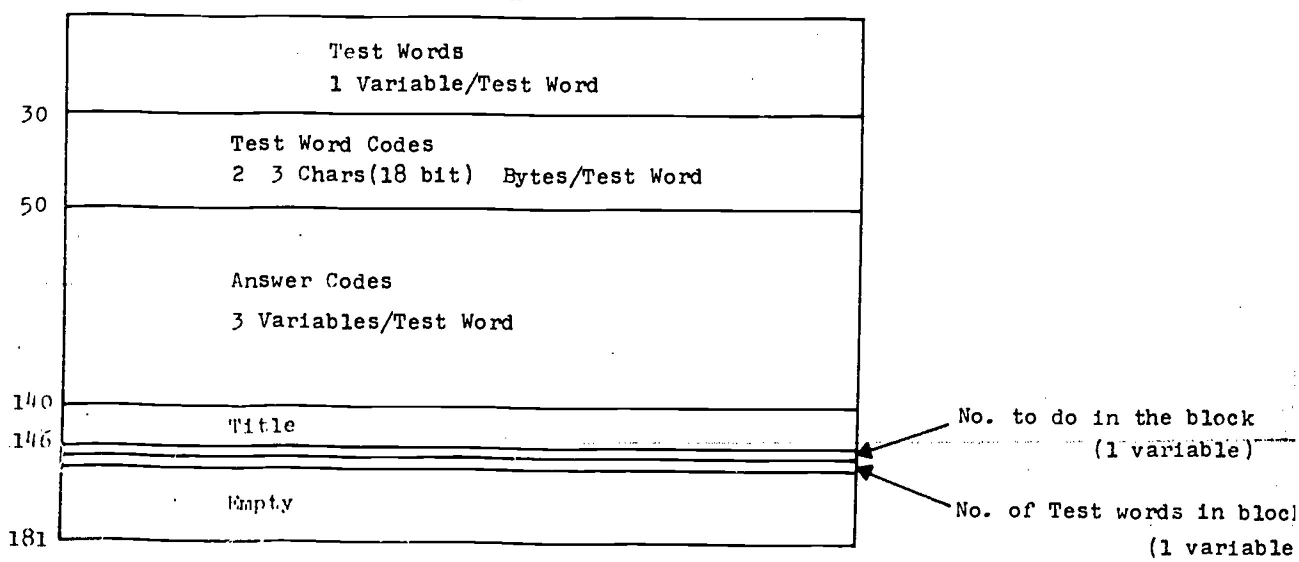


Figure A-2. (cont.)

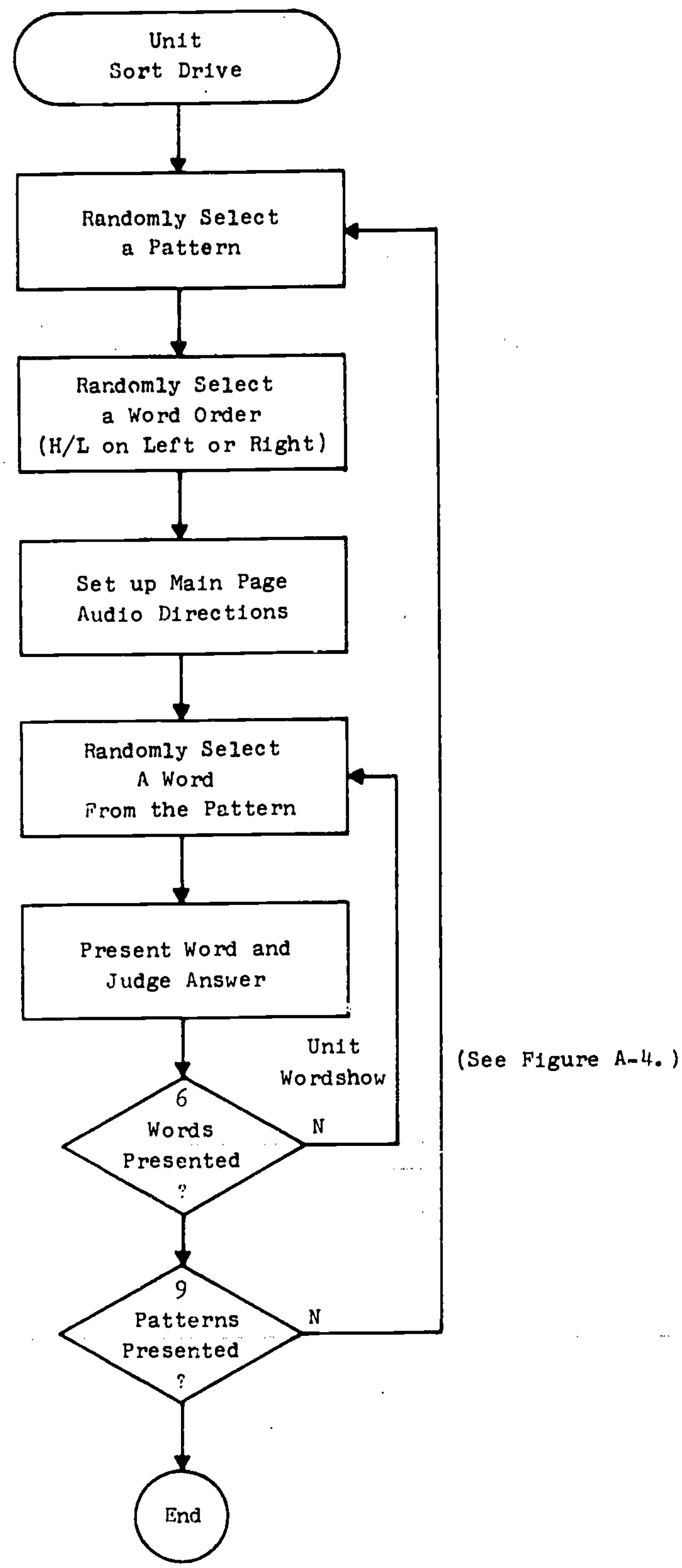


Figure A-3. Flow chart of wordsort test driver.

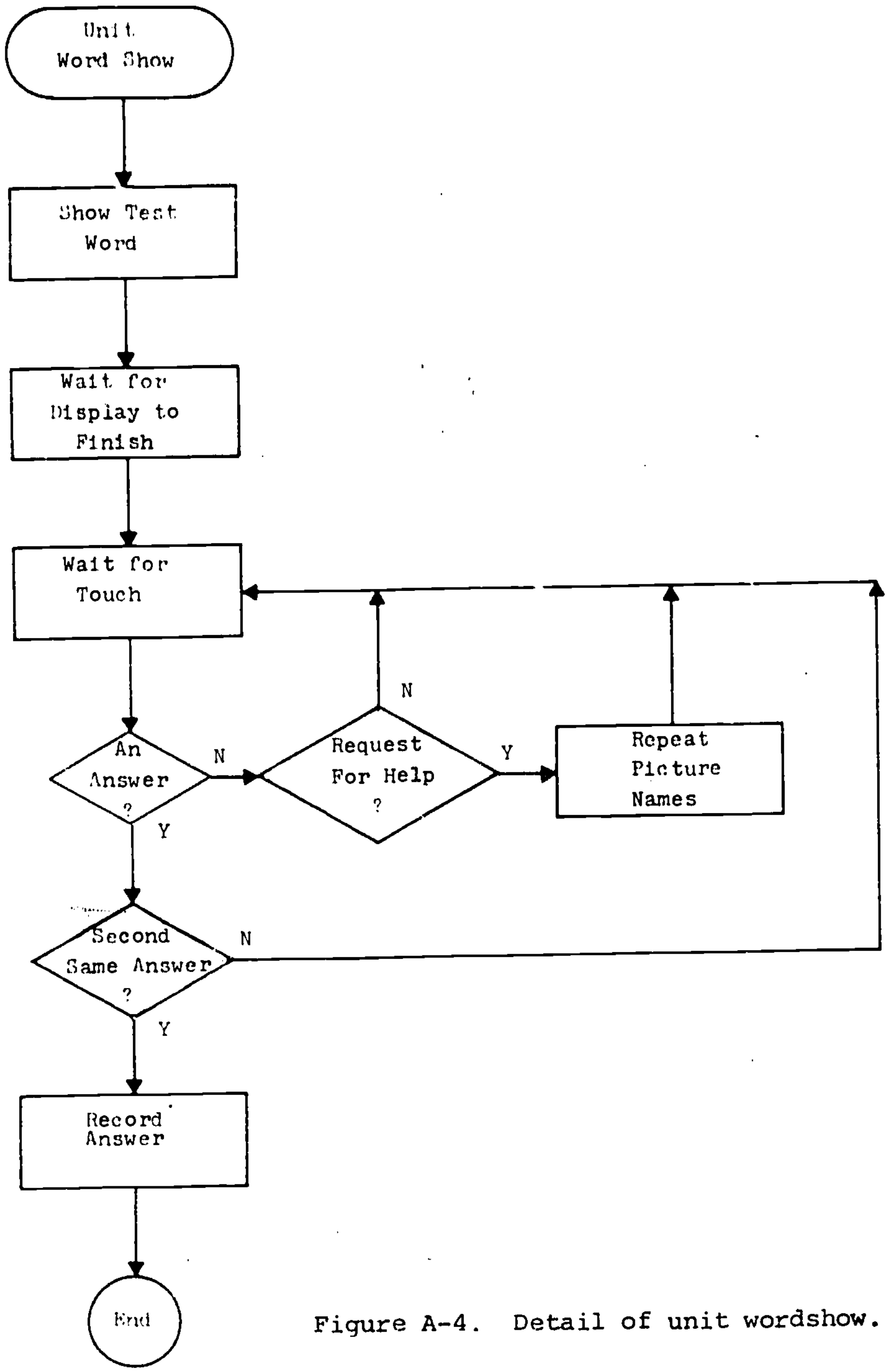


Figure A-4. Detail of unit wordshow.

APPENDIX B

PERCENTAGES CORRECT AND TOTAL NUMBER
OF RESPONSES BY QUARTILES FOR EACH STIMULUS WORD*

*The number under (and slightly to the right of) each percentage score is the total number of responses. Quartiles are formed separately for each pattern.

Initial g - Percent Correct/Total Responses

	Quartile				
Item	Lo	Mid-Lo	Mid-Hi	Hi	All
gest	0/8	0/16	46/24	67/3	25/51
geft	0/5	35/20	48/23	25/4	37/52
gept	0/4	29/17	68/22	67/6	49/49
gim	0/2	35/17	48/21	100/3	44/43
gite	0/5	15/13	33/12	100/4	29/34
geme	0/4	25/12	52/23	67/3	40/42
gind	0/6	8/13	45/29	100/3	33/51
gink	20/5	25/12	50/18	60/5	40/40
gube	14/7	41/17	81/21	100/4	59/49
gurk	25/4	44/9	70/20	100/4	62/37
gon	11/9	43/14	69/26	100/6	56/55
gope	50/2	50/12	56/25	100/6	60/45
gade	0/1	33/9	52/25	100/7	55/42
gabe	0/3	44/16	57/30	86/7	54/56
golb	100/2	60/10	76/21	67/6	72/39
gand	40/5	67/9	65/26	100/7	68/47

Note: Number of Ss in each quartile:
Q1 = 12, Q2 = 36, Q3 = 61, Q4 = 13

Final g - Percent Correct/Total Responses

	Quartile				
Item	Lo	Mid-Lo	Mid-Hi	Hi	All
noge	23/22	21/14	57/14	60/10	37/60
pog	79/14	93/15	84/19	92/12	87/60
bage	33/15	60/15	50/12	87/15	58/57
lig	50/16	48/21	65/20	85/13	60/70
tuge	18/17	31/16	18/11	79/14	36/58
sug	57/14	59/17	92/13	100/12	75/56
darg	14/14	35/17	50/16	83/12	44/59
porge	24/17	33/9	80/10	83/18	56/54
sege	29/14	68/19	58/19	100/14	64/66
mige	17/18	53/17	81/16	100/16	61/67
reg	41/17	44/18	65/17	86/14	58/66
dag	7/14	43/14	92/13	94/18	61/59

Note: Number of Ss in each quartile:
Q1 = 32, Q2 = 32, Q3 = 30, Q4 = 28

Initial c - Percent Correct/Total Responses

			Quartile		
Item	Lo	Mid-Lo	Mid-Hi	Hi	All
cib	0/6	39/23	60/15	67/3	43/47
corp	0/4	28/18	82/11	100/4	49/37
cofe	0/2	52/21	54/13	100/2	53/38
cilf	20/5	12/17	43/7	100/4	30/33
cobe	20/5	79/14	91/11	100/4	76/34
cempt	0/6	35/20	71/7	80/5	42/38
cung	0/2	55/22	83/12	83/6	64/42
cyfe	0/4	35/20	67/6	67/6	42/36
cabe	50/2	62/13	75/8	100/3	69/26
cefe	40/5	35/20	44/9	67/6	43/40
cest	0/4	0/17	40/10	100/8	31/39
carg	83/6	50/22	69/13	100/5	65/46
cylm	33/3	50/14	57/14	80/5	56/36
cuse	0/2	47/19	69/13	100/4	58/38
cipe	0/2	50/22	64/11	100/7	60/42
cirt	0/6	27/11	44/9	86/7	39/33
cym	0/0	36/14	75/8	100/1	52/23
cose	50/2	55/22	76/17	88/8	67/49
cade	0/3	33/12	100/6	86/7	57/28
corb	0/3	62/13	75/4	71/7	59/27

Note: Number of Ss in each quartile:
Q1 = 12, Q2 = 59, Q3 = 34, Q4 = 17

Final c - Percent Correct/Total Responses

		Quartile			
Item	Lo	Mid-Lo	Mid-Hi	Hi	All
woc	60/15	17/6	100/15	88/17	75/53
dac	37/19	29/14	57/14	91/11	50/58
jic	36/14	50/16	71/14	88/16	62/60
roce	6/18	47/15	47/15	73/11	39/59
bec	21/14	67/15	85/13	100/14	68/56
nuc	20/15	58/12	90/10	100/14	65/51
hac	28/25	44/16	50/12	94/16	51/69
fice	11/18	50/16	70/20	77/13	51/67
mic	20/20	39/18	50/14	100/15	49/67
tuce	35/20	69/16	53/15	83/12	57/63
poce	16/19	43/14	62/13	100/10	48/56
mece	37/19	64/22	69/13	79/19	62/73

Note: Number of Ss in each quartile:
Q1 = 36, Q2 = 30, Q3 = 28, Q4 = 28

Monosyllabic a - Percent Correct/Total Responses

Quartile

Item	Lo	Mid-Lo	Mid-Hi	Hi	All
cale	30/10	50/8	83/6	94/17	68/41
dack	33/9	33/6	88/8	100/7	63/30
lase	25/8	36/14	25/4	86/7	42/33
tasp	33/6	57/7	88/8	86/7	68/28
dape	30/10	50/8	33/6	100/11	57/35
prane	0/2	25/8	50/2	100/8	55/20
bab	75/4	60/5	78/9	100/11	83/29
vabe	0/6	67/3	71/7	92/12	64/28
fam	50/8	100/4	63/8	100/9	76/29
nad	0/4	50/6	67/6	89/9	60/25
tafe	50/6	29/7	43/7	92/12	59/32
tand	0/9	63/8	89/9	78/9	57/35
plab	0/8	43/7	67/6	94/17	61/38
tass	75/8	50/6	71/7	92/12	76/33
plafe	25/8	57/7	100/4	92/12	68/31
bap	0/2	100/3	86/7	100/6	83/18
hade	0/3	40/5	50/6	82/11	56/25
lan	25/8	60/10	60/5	100/11	65/34
slape	33/9	17/6	57/7	93/15	59/37
dap	29/7	71/7	89/9	100/6	72/29
trabe	29/7	57/7	33/3	100/11	64/28
nast	40/5	71/7	83/6	100/10	79/28
calge	0/3	38/8	25/8	73/11	43/30
trabe*	17/6	45/11	50/8	91/11	56/36

Note: Number of Ss in each quartile:
Q1 = 26, Q2 = 28, Q3 = 26, Q4 = 42

*Included twice by mistake.

Bisyllabic a - Percent Correct/Total Responses

Quartile

Item	Lo	Mid-Lo	Mid-Hi	Hi	All
nattis	26/19	58/19	55/11	60/10	47/59
dallor	38/21	55/22	75/12	100/9	59/64
nammis	42/19	56/25	55/11	83/6	54/61
lapple	40/15	65/17	92/12	75/8	65/52
cannod	17/18	45/22	81/16	83/6	50/62
sattom	53/17	80/15	81/16	100/8	75/56
natis	29/24	30/20	56/18	67/9	41/71
dalor	22/23	55/20	53/17	100/8	49/68
namis	20/20	55/20	58/19	89/9	50/68
laple	33/21	44/16	93/15	100/7	59/59
canod	10/21	29/17	50/12	100/6	34/56
satom	13/16	33/21	53/15	75/4	36/56

Note: Number of Ss in each quartile:
Q1 = 39, Q2 = 39, Q3 = 29, Q4 = 15

Monosyllabic i - Percent Correct/Total Responses

Quartile

Item	Lo	Mid-Lo	Mid-Hi	Hi	All
jik	46/26	60/10	71/14	100/15	66/65
slin	35/20	50/12	80/10	100/18	65/60
kip	47/19	63/8	79/14	100/18	73/59
bim	37/19	89/9	100/13	100/16	77/57
jit	45/22	58/19	93/14	100/16	70/71
sib	27/15	22/9	92/12	100/16	63/52
jike	36/22	77/13	93/15	100/16	73/66
sline	59/17	75/12	82/11	100/17	79/57
kipe	24/21	74/19	78/9	100/14	63/63
bime	42/26	80/10	100/8	100/13	70/57
jite	33/24	77/13	67/12	100/15	64/64
sibe	33/21	70/10	67/12	100/18	66/61

Note: Number of Ss in each quartile:
Q1 = 42, Q2 = 24, Q3 = 24, Q4 = 32

Monosyllabic o - Percent Correct/Total Responses

Quartile

Item	Lo	Mid-Lo	Mid-Hi	Hi	All
bod	20/5	57/14	56/16	100/33	75/68
mot	0/5	78/18	85/13	96/27	81/63
wom	33/9	80/15	71/21	96/28	78/73
tob	29/7	33/15	58/19	89/19	58/60
rog	25/8	67/9	86/14	96/24	78/55
lop	33/6	40/15	94/18	100/28	79/67
mote	44/9	44/9	46/13	91/23	65/54
wome	33/9	27/15	67/15	96/26	65/65
tobe	0/9	46/13	47/17	92/26	58/65
roge	33/6	38/13	50/12	91/23	63/54
lope	33/6	30/10	76/17	95/19	69/52
jode	40/5	50/16	55/11	96/24	70/56

Note: Number of Ss in each quartile:
Q1 = 14, Q2 = 27, Q3 = 31, Q4 = 50

Monosyllabic u - Percent Correct/Total Responses

		Quartile			
Item	Lo	Mid-Lo	Mid-Hi	Hi	All
fub	33/12	50/12	86/14	100/21	73/59
pud	50/12	77/13	71/17	100/19	77/61
vun	43/14	69/16	100/9	100/22	79/61
huk	25/16	63/16	76/17	89/19	65/68
glub	46/13	78/9	69/13	95/21	75/56
stut	38/21	46/13	77/13	93/14	61/61
fube	31/16	33/18	40/15	81/16	46/65
pude	25/16	22/9	41/17	85/13	44/55
vune	19/16	38/16	91/11	88/16	56/59
huke	12/17	40/15	43/14	89/18	47/64
glube	21/19	38/16	69/13	87/15	51/63
stute	29/14	56/9	60/15	100/22	67/60

Note: Number of Ss in each quartile:
Q1 = 31, Q2 = 27, Q3 = 28, Q4 = 36